

NCBI Molecular Biology Resources

Part 2: Using NCBI BLAST

December 2009

NCBI
Public Services

Using BLAST

- Basics of using NCBI BLAST
- Using the new Interface
 - Improved organism and filter options
- New Services
 - Primer BLAST
 - Align 2 Sequences Integration
 - COBALT – protein multiple alignment
- BLAST URL API
- C++ BLAST binaries

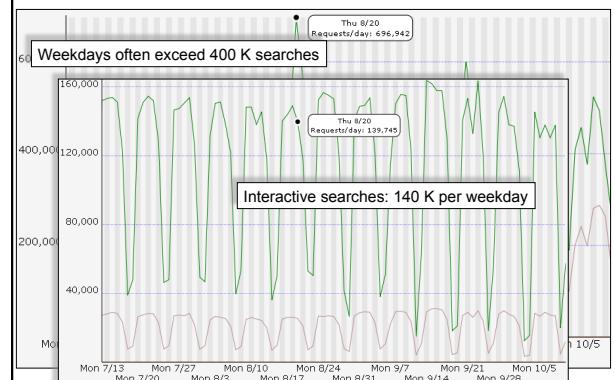
NCBI
Public Services

Basic Local Alignment Search Tool

- Widely used similarity search tool
- Heuristic approach based on Smith Waterman algorithm
- Finds best local alignments
- Provides statistical significance
- All combinations (DNA/Protein) query and database.
 - DNA vs DNA
 - DNA translation vs Protein
 - Protein vs Protein
 - Protein vs DNA translation
 - DNA translation vs DNA translation
- www, standalone, and network client

NCBI
Public Services

BLAST Activity

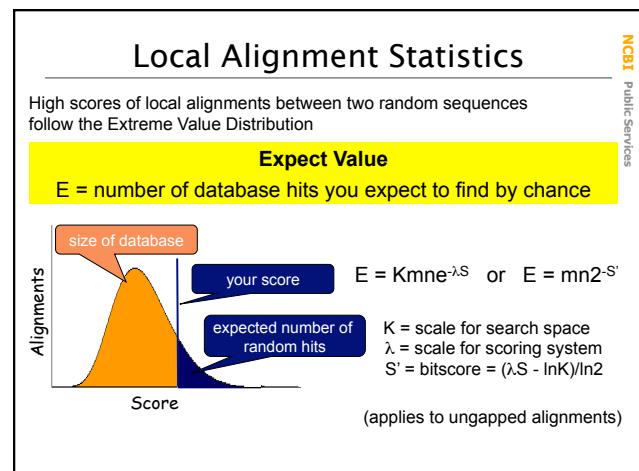
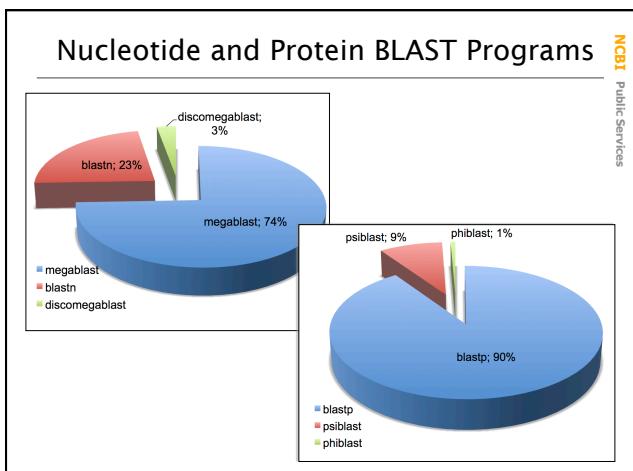
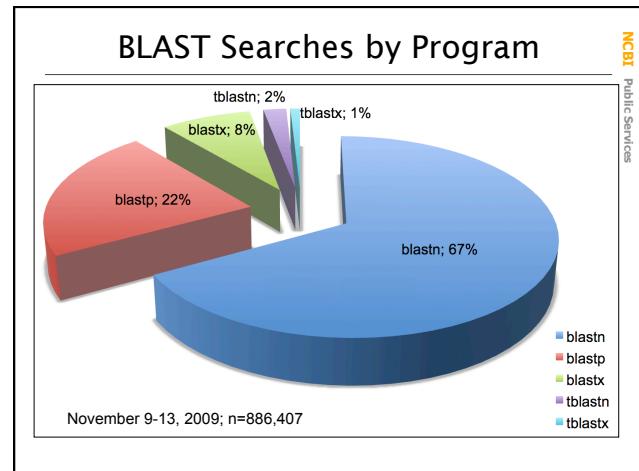


NCBI
Public Services

BLAST and BLAST-like programs

- Traditional BLAST (formerly blastall) nucleotide, protein, translations
 - blastn nucleotide query vs. nucleotide database
 - blastp protein query vs. protein database
 - blastx nucleotide query vs. protein database
 - tblastn protein query vs. translated nucleotide database
 - tblastx translated query vs. translated database
- Megablast nucleotide only
 - Contiguous megablast
 - Nearly identical sequences
 - Discontiguous megablast
 - Cross-species comparison
- Position Specific BLAST Programs protein only
 - Position Specific Iterative BLAST (PSI-BLAST)
 - Automatically generates a position specific score matrix (PSSM)
 - Reverse PSI-BLAST (RPS-BLAST)
 - Searches a database of PSI-BLAST PSSMs

NCBI Public Services



The BLAST homepage

<http://blast.ncbi.nlm.nih.gov/>

NCBI Public Services

Basic BLAST: Databases

Non-redundant protein

Choose Search Set

Database	Description
nr (non-redundant protein sequences)	- GenBank CDS translations - NP_, XP_, refseq_protein - Outside Protein • PIR, Swiss-Prot, PRF • PDB (sequences from structures)
pat	protein patents
env_nr	environmental samples

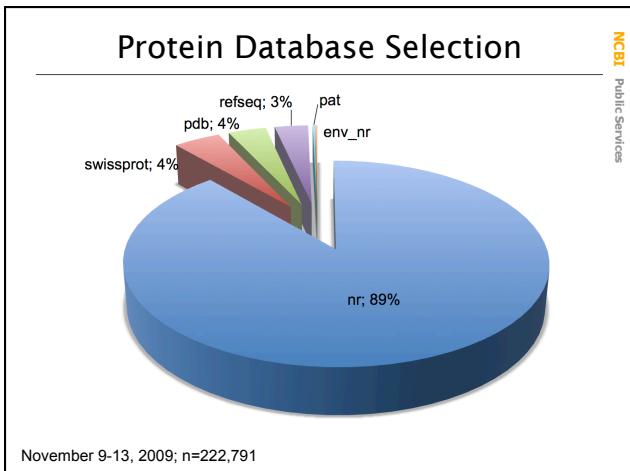
Services
blastp
blastx

NCBI Public Services

Protein Database Sizes

Database	Sequences	Residues
nr	10,133,783	3,456,922,644
refseq_protein	7,413,069	2,589,005,568
swissprot	430,511	159,291,105
pat	817,680	166,184,433
pdb	44,202	10,171,945

12/04/2009



Nucleotide Databases: Human and Mouse

Megablast, blastn service

Choose Search Set

Database Human genomic + transcript Mouse genomic + transcript Others (nr etc.)
 Human genomic plus transcript

- Human and mouse genomic and transcript now default
- Separate sections in output for mRNA and genomic
- Direct links to Map Viewer for genomic sequences

Nucleotide Databases: Traditional

Choose Search Set

Database Human genomic + transcript Mouse genomic + transcript Others (nr etc.)

Organism **Optional**

Exclude **Optional**

Entrez Query **Optional**

Other Databases

Nucleotide collection (nr/nt)

- Reference mRNA sequences (refseq_rna)
- Reference genomic sequences (refseq_genomic)
- NCBI Genomes (chromosome)
- Expressed sequence tags (est)
- Non-human, non-mouse ESTs (est_others)
- Genomic survey sequences (gss)
- High throughput genomic sequences (HTGS)
- Patent sequences (pat)
- Protein Data Bank (pdb)
- Human ALU repeat elements (alu_repeats)
- Sequence tagged sites (dbsts)
- Whole-genome shotgun reads (wgs)
- Environmental samples (env_nt)

Services

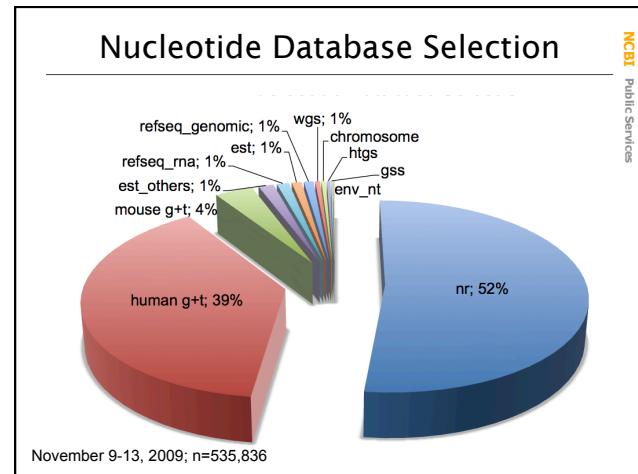
blastn
tblastn
tblastx

- ## Nucleotide Databases: Traditional
- Databases are mostly non-overlapping
- **nr (nt)**
 - Traditional GenBank
 - NM_ and XM_ RefSeqs
 - refseq_rna
 - **NCBI Genomes**
 - NC_RefSeqs
 - GenBank Chromosomes
 - **dbest**
 - EST Division
 - non-human, non-mouse ests
 - **htgs**
 - HTG division
 - **gss**
 - GSS division
 - **wgs**
 - whole genome shotgun contigs
 - **env_nt**
 - environmental samples

Nucleotide Database Sizes

NCBI Public Services
12/04/2009

Database	Sequences	Residues
nr/nt	10,362,162	29,617,088,643
refseq_rna	2,042,538	3,240,301,155
NCBI genomes	10,047	49,094,451,709
est	63,832,451	35,136,825,005
htgs	143,742	24,082,224,044
gss	27,198,629	17,658,377,015
wgs	31,377,631	149,309,157,200
env_nt	17,708,548	7,218,208,433



Using Basic BLAST

NCBI Public Services

Universal Form: Protein

NCBI Public Services

Program Selection

blastp (protein-protein BLAST)
 PSI-BLAST (Position-Specific Iterated BLAST)
 PHI-BLAST (Pattern Hit Initiated BLAST)

Choose a BLAST algorithm

Algorithm parameters

Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)

Show results in a new window

Universal Form: Nucleotide

NCBI Public Services

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

Program Selection

Optimize for Sensitivity Speed

Less More

Highly similar sequences (megablast)
 More dissimilar sequences (discontiguous megablast)
 Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Less More

Highly similar sequences (megablast)
 More dissimilar sequences (discontiguous megablast)
 Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST Search database Human G+T using Megablast (Optimize for highly similar sequences) Show results in a new window

Limiting Database: Organism

NCBI Public Services

Choose Search Set

Database Human genomic + transcript Mouse genomic + transcript Others (nr etc.)

Organism Optional

Exclude Optional

Entrez Query Optional

mammal Exclude 20 top taxa will be shown.

bacteria

Organism autocomplete

mammals (taxid:40674)
Mammalia (taxid:40674)
eutherian mammals (taxid:9347)
placental mammals (taxid:9347)
egg-laying mammals (taxid:9255)
mammalian hepatitis B-type viruses (taxid:10405)
"Streptococcus de la mammis" Nocard and Mollereau 1894
Mammalian virus group (taxid:353212)
Mouse mammary tumor virus (taxid:11757)
Mammuthus (taxid:37348)

Combining Organisms

NCBI Public Services

Choose Search

Primates and Rodents without human or mouse

Database Human genomic + transcript Mouse genomic + transcript Others (nr etc.)

Nucleotide collection (nr/nt)

Organism Optional

primates (taxid:9443) Exclude
human (taxid:9606) Exclude
rodents (taxid:9989) Exclude
mouse (taxid:10090) Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

More Limits

NCBI Public Services

Eliminate models and environmental samples

Exclude Models (XM/XP) Sequences from uncultured bacteria

Entrez Query Optional

biomol_mrna[Prop] AND srcdb_ddbj/embl/genbank[Prop]
Enter an Entrez query to limit search

Entrez query limit, any valid Entrez query.

Algorithm parameters: Protein

General Parameters

- Max target sequences: 100 (selected)
- Short queries: Automatically adjust parameters for short queries
- Expect threshold: 10 (selected)
- Word size: 3 (selected)

Scoring Parameters

- Matrix: BLOSUM62
- Gap Costs: Existence: 11 Extension: 1
- Compositional adjustments: Conditional compositional score matrix adjustment

Filters and Masking

- Filter: Low complexity regions
- Mask: Mask for lookup table only, Mask lower case letters

Expand button (highlighted) leads to a dropdown menu with options: Max target sequences, Short queries, Expect threshold, Word size, Scoring Parameters, Filters and Masking.

May limit results (highlighted) points to the 'Automatically adjust parameters for short queries' checkbox.

Adjust to set stringency (highlighted) points to the 'Expect threshold' input field.

Default statistics adjustment for compositional bias (highlighted) points to the 'Conditional compositional score matrix adjustment' checkbox.

Off now by default. Conflicts with comp-based stats (highlighted) points to the 'Mask for lookup table only' checkbox.

Automatic Short Sequence Adjustment

Job Title: Elvis Lives!

No putative conserved domains have been detected

Your search parameters were adjusted to search for a short input sequence.

Request ID: HGC9100G
Status: Searching
Submitted at: Thu Dec 3 18:53 (EST)
Current time: Thu Dec 3 18:53 (EST)
Time since submission: 00:00:08

This page will be automatically updated in 5 seconds

Protein

e-value	20000
Word Size	2
Matrix	PAM30
Comp Stats	Off
Low Comp Filter	Off

Nucleotide

e-value	1000
Word Size	7
Matrix	1,-3
Low Comp Filter	Off

New Designing or Testing PCR Primers? Try your search in Primer-BLAST. [Go](#)

Algorithm parameters: Nucleotide

General Parameters

- Max target sequences: 100 (selected)
- Short queries: Automatically adjust parameters for short queries
- Expect threshold: 10 (selected)
- Word size: 11 (selected)

Scoring Parameters

- Match/Mismatch: 2,-3
- Scores: 5
- Gap Costs: Existence: 5 Extension: 2

Filters and Masking

- Filter: Low complexity regions, Species-specific repeats for: Human
- Mask: Mask for lookup table only, Mask lower case letters

blastn button (highlighted) leads to a dropdown menu with options: Max target sequences, Short queries, Expect threshold, Word size, Scoring Parameters, Filters and Masking.

Masks species-specific interspersed repeats (highlighted) points to the 'Species-specific repeats for: Human' dropdown menu.

Essential for genomic query sequences (highlighted) points to the 'Species-specific repeats for: Human' dropdown menu.

Masks LC sequence (simple repeats) (highlighted) points to the 'Mask for lookup table only' checkbox.

Prevents starting alignment in masked region (highlighted) points to the 'Mask for lookup table only' checkbox.

Allows extensions through masked regions (highlighted) points to the 'Mask for lookup table only' checkbox.

Basic BLAST: Protein

The hard way to run a BLAST Search

The hard way to run a BLAST Search

1. Search protein with "Human Muscle Creatine Kinase"

2. Click on summary for NP_001815

3. Change format to FASTA

4. Select sequence

5. Copy sequence

6. Google search "BLAST"

7. Link to NCBI BLAST Homepage

8. Link to Protein BLAST form

9. Paste FASTA sequence into form

10. Click BLAST button

muscle creatine kinase [Homo sapiens]
1. 381 aa protein
NP_001815.2 GI:21536288

muscle creatine kinase [Homo sapiens]
NP_001815.2 GI:21536288

NCBI B

ncbi blast Suggestions

ncbi books

ncbi blast download

ncbi blast2

ncbi blast tutorial

ncbi biosystems

protein blast

Search protein database using a protein query
Algorithms: blastp, psi-blast, phi-blast

Enter accession number, gi, or FASTA sequence

>gi|21536288|ref|NP_001815.2| muscle creatine kinase [Homo sapiens]

MPPGPNHKKFLKNNYKPEEYYPDLSKHNHHMVAU...TLEYLKLDRKEPTFSPVWDIVOTGVDPNPHP
TVGGCVAGDEESVYKEELPFIISDRGKGYKPTFDLNLHNLNEKLGDOLDPNVPYSLSVKRVKG
WCPBFLRSRGERRAVKRLSREALNSLTGEFKQYIPLKTSWMLTGTGKQYIHDHFLDFPVPSPLASGM
WCPARGIHNNDKGFSLVWNNEEDHLRN13MEKCGNGMKVTRFCVGQLQH13EII1T1KAGHFPYMNQH
GVY

Clear

BLAST

An easier way: Entrez protein record

Format: GenPep FASTA Gisapca More Formats ▾

Try the [GraphDB](#) report for a more informative view of the biological features.

NCBI Reference Sequence: NP_001912

muscle creatine kinase [homo sapiens]

[Comment](#) [Feature](#) [Sequence](#)

Locus: NP_001912 S38 as linear PM 24-JUL-2009

DEFINITION muscle creatine kinase [homo sapiens].
ACCESSION NP_001912
VERSION NP_001912.5
SOURCE gi|121753288
ORGANISM Homo sapiens (human)
ORGANISM Homo sapiens (human)

METADATA (Metasource: Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Boreoeutheria; Primates; Haplorhini; Hominoidea; Hominidae; H. sapiens)

REFERENCES

AUTHORS Santopietro,C., Zelante,L., Minneci,A., Gonzales-Freire,M., Puccetti,P., Cicali,C., ... De Luca,G.
TITLE Does the polymeric protein determine the potential for becoming a biomarker of exercise-induced muscle damage? The effect of rowing
JOURNAL Scand J Med Sci Sports 19 (in press)

JOURNAL Scand J Med Sci Sports 19 (in press)

REMARK GenBank: Observational study of gene-disease association and expression analysis of the human creatine kinase gene.

REFERENCE

AUTHORS Kian, H.-D., ...
TITLE Low-carbohydrate diet reduces serum creatine kinase (CK) levels and CPK and plasma levels to running distance in a marathon runner. A case report. Int J Appl Physiol. 105 (5), 765-770 (2009)

JOURNAL Int J Appl Physiol. 105 (5), 765-770 (2009)

REMARK

REMARK In this study, the serum CK and CPK levels increased up to threshold at the end of the race. This was further increased after the race. The serum CK and CPK levels were reduced by 3-fold at the end of the 200-km race.

REFERENCE

AUTHORS Vilaneu,K., Anderson,J., Heggelund,T., Farstad,I., and Skjeldal,O.H.
TITLE Low-probability value of creatine kinase-isozymic pattern in blood samples from patients with suspected myocardial infarction. Clin Cardiol 13 (2), 122-125 (2000)

JOURNAL Clin Cardiol 13 (2), 122-125 (2000)

REMARK GenBank: CK-MB values < 5 times the reference after exercise may indicate myocardial damage.

REFERENCE

AUTHORS Madsen, M., ...
TITLE Creatine kinase isoenzymes in healthy individuals. Clin Biochem 1 (1 to 381)

JOURNAL Clin Biochem 1 (1 to 381)

REMARK GenBank: CK-MM values < 5 times the reference after exercise may indicate myocardial damage.

Change Region Shown ▾

Custom View ▾

Sequence Analysis Tools

BLAST Sequence
Conserved Domains
Articles about the CKM gene

Long-term prognostic value
Unpublished data
Duchenne muscular dystrophy
ideopathic [PubMed] (Last J Paediatr Neurol 2006)

Significance of CK elevation
nomination [PubMed] (Last J Cardiol 2006)

See all ▾

PubMed Citations

Identical Proteins

Identical Protein for NP_001912

Sequence 8888 from penter
creative kinase, muscle [Nr_AASTR03]
Chen, D. Crystal Structure of HcK. BiJ 2005

RefSeq in mRNA

See reference in RNA sequence
CKM gene (NM_001912)

Reference Sequences

More about the CKM gene

The CKM gene encodes creatine kinase, muscle, which is a cytosolic enzyme involved in energy metabolism and a useful marker for myocardial enzyme.

Also Known As CKMM, HcK

Homologs of the CKM gene

The CKM gene is conserved in chimpanzee, rhesus monkey, and bat.

Gene Record

HomoloGene Cluster

BLAST Ad to BLAST form

BLAST Ad to BLAST form

muscle creatine kinase [Homo sapiens]

[Customize View](#)

Comment
Features
Sequence

[Sequence Analysis Tools](#)

LOCUS	NP_001815	381 aa	linear	PRI	26-JUL-2004	 BLAST Sequence
DEFINITION	muscle creatine kinase [Homo sapiens].					
ACCESSION	NP_001815					

BLASTP programs search protein databases using a protein query. [more...](#)

Enter Query Sequence

Clear

Query subrange

NP_001815.2

From

To

Or, upload file

Browse...

NP_001815:muscle creatine kinase [Homo sapiens]

Enter a descriptive title for your BLAST search
 (optional)

Align two or more sequences
 (optional)

Database and limits

Database and limits

Choose Search Set

Database	<input type="text" value="Reference proteins (refseq_protein)"/>	NCBI Reference Sequences
Organism	<input type="checkbox"/> mammals (taxid:40674) <input type="checkbox"/> primate (taxid:9443)	
Mammals without primates	<input type="checkbox"/> Exclude <input checked="" type="checkbox"/> Exclude	
Exclude	Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.	
Optional	<input checked="" type="checkbox"/> Models (XM/XP) <input type="checkbox"/> Environmental sample sequences	
Entrez Query	<input type="text" value="Exclude predicted proteins"/> <input type="text" value="Enter an Entrez query to limit search"/>	
Optional		

Run Search

Program Selection

Algorithm

- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)

Choose a BLAST algorithm

BLAST

Search database Reference proteins (refseq_protein) using Blastp (protein-protein BLAST)
 Show results in a new window

NCBI
Public Services

BLAST Formatting Page

NCBI BLAST/blastp suite/ Formatting Results - HGEZFJEN012 [Formatting options]

Job Title: gj|21536288|ref|NP_001615.2| muscle creatine...

Putative conserved domains have been detected, click on the image below for detailed results.

Query seq. superfamily 1 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000 1050 1100 1150 1200 1250 1300 1350 1400 1450 1500 1550 1600 1650 1700 1750 1800 1850 1900 1950 2000 2050 2100 2150 2200 2250 2300 2350 2400 2450 2500 2550 2600 2650 2700 2750 2800 2850 2900 2950 3000 3050 3100 3150 3200 3250 3300 3350 3400 3450 3500 3550 3600 3650 3700 3750 3800 3850 3900 3950 4000 4050 4100 4150 4200 4250 4300 4350 4400 4450 4500 4550 4600 4650 4700 4750 4800 4850 4900 4950 5000 5050 5100 5150 5200 5250 5300 5350 5400 5450 5500 5550 5600 5650 5700 5750 5800 5850 5900 5950 6000 6050 6100 6150 6200 6250 6300 6350 6400 6450 6500 6550 6600 6650 6700 6750 6800 6850 6900 6950 7000 7050 7100 7150 7200 7250 7300 7350 7400 7450 7500 7550 7600 7650 7700 7750 7800 7850 7900 7950 8000 8050 8100 8150 8200 8250 8300 8350 8400 8450 8500 8550 8600 8650 8700 8750 8800 8850 8900 8950 9000 9050 9100 9150 9200 9250 9300 9350 9400 9450 9500 9550 9600 9650 9700 9750 9800 9850 9900 9950 10000 10050 10100 10150 10200 10250 10300 10350 10400 10450 10500 10550 10600 10650 10700 10750 10800 10850 10900 10950 11000 11050 11100 11150 11200 11250 11300 11350 11400 11450 11500 11550 11600 11650 11700 11750 11800 11850 11900 11950 12000 12050 12100 12150 12200 12250 12300 12350 12400 12450 12500 12550 12600 12650 12700 12750 12800 12850 12900 12950 13000 13050 13100 13150 13200 13250 13300 13350 13400 13450 13500 13550 13600 13650 13700 13750 13800 13850 13900 13950 14000 14050 14100 14150 14200 14250 14300 14350 14400 14450 14500 14550 14600 14650 14700 14750 14800 14850 14900 14950 15000 15050 15100 15150 15200 15250 15300 15350 15400 15450 15500 15550 15600 15650 15700 15750 15800 15850 15900 15950 16000 16050 16100 16150 16200 16250 16300 16350 16400 16450 16500 16550 16600 16650 16700 16750 16800 16850 16900 16950 17000 17050 17100 17150 17200 17250 17300 17350 17400 17450 17500 17550 17600 17650 17700 17750 17800 17850 17900 17950 18000 18050 18100 18150 18200 18250 18300 18350 18400 18450 18500 18550 18600 18650 18700 18750 18800 18850 18900 18950 19000 19050 19100 19150 19200 19250 19300 19350 19400 19450 19500 19550 19600 19650 19700 19750 19800 19850 19900 19950 20000 20050 20100 20150 20200 20250 20300 20350 20400 20450 20500 20550 20600 20650 20700 20750 20800 20850 20900 20950 21000 21050 21100 21150 21200 21250 21300 21350 21400 21450 21500 21550 21600 21650 21700 21750 21800 21850 21900 21950 22000 22050 22100 22150 22200 22250 22300 22350 22400 22450 22500 22550 22600 22650 22700 22750 22800 22850 22900 22950 23000 23050 23100 23150 23200 23250 23300 23350 23400 23450 23500 23550 23600 23650 23700 23750 23800 23850 23900 23950 24000 24050 24100 24150 24200 24250 24300 24350 24400 24450 24500 24550 24600 24650 24700 24750 24800 24850 24900 24950 25000 25050 25100 25150 25200 25250 25300 25350 25400 25450 25500 25550 25600 25650 25700 25750 25800 25850 25900 25950 26000 26050 26100 26150 26200 26250 26300 26350 26400 26450 26500 26550 26600 26650 26700 26750 26800 26850 26900 26950 27000 27050 27100 27150 27200 27250 27300 27350 27400 27450 27500 27550 27600 27650 27700 27750 27800 27850 27900 27950 28000 28050 28100 28150 28200 28250 28300 28350 28400 28450 28500 28550 28600 28650 28700 28750 28800 28850 28900 28950 29000 29050 29100 29150 29200 29250 29300 29350 29400 29450 29500 29550 29600 29650 29700 29750 29800 29850 29900 29950 30000 30050 30100 30150 30200 30250 30300 30350 30400 30450 30500 30550 30600 30650 30700 30750 30800 30850 30900 30950 31000 31050 31100 31150 31200 31250 31300 31350 31400 31450 31500 31550 31600 31650 31700 31750 31800 31850 31900 31950 32000 32050 32100 32150 32200 32250 32300 32350 32400 32450 32500 32550 32600 32650 32700 32750 32800 32850 32900 32950 33000 33050 33100 33150 33200 33250 33300 33350 33400 33450 33500 33550 33600 33650 33700 33750 33800 33850 33900 33950 34000 34050 34100 34150 34200 34250 34300 34350 34400 34450 34500 34550 34600 34650 34700 34750 34800 34850 34900 34950 35000 35050 35100 35150 35200 35250 35300 35350 35400 35450 35500 35550 35600 35650 35700 35750 35800 35850 35900 35950 36000 36050 36100 36150 36200 36250 36300 36350 36400 36450 36500 36550 36600 36650 36700 36750 36800 36850 36900 36950 37000 37050 37100 37150 37200 37250 37300 37350 37400 37450 37500 37550 37600 37650 37700 37750 37800 37850 37900 37950 38000 38050 38100 38150 38200 38250 38300 38350 38400 38450 38500 38550 38600 38650 38700 38750 38800 38850 38900 38950 39000 39050 39100 39150 39200 39250 39300 39350 39400 39450 39500 39550 39600 39650 39700 39750 39800 39850 39900 39950 40000 40050 40100 40150 40200 40250 40300 40350 40400 40450 40500 40550 40600 40650 40700 40750 40800 40850 40900 40950 41000 41050 41100 41150 41200 41250 41300 41350 41400 41450 41500 41550 41600 41650 41700 41750 41800 41850 41900 41950 42000 42050 42100 42150 42200 42250 42300 42350 42400 42450 42500 42550 42600 42650 42700 42750 42800 42850 42900 42950 43000 43050 43100 43150 43200 43250 43300 43350 43400 43450 43500 43550 43600 43650 43700 43750 43800 43850 43900 43950 44000 44050 44100 44150 44200 44250 44300 44350 44400 44450 44500 44550 44600 44650 44700 44750 44800 44850 44900 44950 45000 45050 45100 45150 45200 45250 45300 45350 45400 45450 45500 45550 45600 45650 45700 45750 45800 45850 45900 45950 46000 46050 46100 46150 46200 46250 46300 46350 46400 46450 46500 46550 46600 46650 46700 46750 46800 46850 46900 46950 47000 47050 47100 47150 47200 47250 47300 47350 47400 47450 47500 47550 47600 47650 47700 47750 47800 47850 47900 47950 48000 48050 48100 48150 48200 48250 48300 48350 48400 48450 48500 48550 48600 48650 48700 48750 48800 48850 48900 48950 49000 49050 49100 49150 49200 49250 49300 49350 49400 49450 49500 49550 49600 49650 49700 49750 49800 49850 49900 49950 50000 50050 50100 50150 50200 50250 50300 50350 50400 50450 50500 50550 50600 50650 50700 50750 50800 50850 50900 50950 51000 51050 51100 51150 51200 51250 51300 51350 51400 51450 51500 51550 51600 51650 51700 51750 51800 51850 51900 51950 52000 52050 52100 52150 52200 52250 52300 52350 52400 52450 52500 52550 52600 52650 52700 52750 52800 52850 52900 52950 53000 53050 53100 53150 53200 53250 53300 53350 53400 53450 53500 53550 53600 53650 53700 53750 53800 53850 53900 53950 54000 54050 54100 54150 54200 54250 54300 54350 54400 54450 54500 54550 54600 54650 54700 54750 54800 54850 54900 54950 55000 55050 55100 55150 55200 55250 55300 55350 55400 55450 55500 55550 55600 55650 55700 55750 55800 55850 55900 55950 56000 56050 56100 56150 56200 56250 56300 56350 56400 56450 56500 56550 56600 56650 56700 56750 56800 56850 56900 56950 57000 57050 57100 57150 57200 57250 57300 57350 57400 57450 57500 57550 57600 57650 57700 57750 57800 57850 57900 57950 58000 58050 58100 58150 58200 58250 58300 58350 58400 58450 58500 58550 58600 58650 58700 58750 58800 58850 58900 58950 59000 59050 59100 59150 59200 59250 59300 59350 59400 59450 59500 59550 59600 59650 59700 59750 59800 59850 59900 59950 60000 60050 60100 60150 60200 60250 60300 60350 60400 60450 60500 60550 60600 60650 60700 60750 60800 60850 60900 60950 61000 61050 61100 61150 61200 61250 61300 61350 61400 61450 61500 61550 61600 61650 61700 61750 61800 61850 61900 61950 62000 62050 62100 62150 62200 62250 62300 62350 62400 62450 62500 62550 62600 62650 62700 62750 62800 62850 62900 62950 63000 63050 63100 63150 63200 63250 63300 63350 63400 63450 63500 63550 63600 63650 63700 63750 63800 63850 63900 63950 64000 64050 64100 64150 64200 64250 64300 64350 64400 64450 64500 64550 64600 64650 64700 64750 64800 64850 64900 64950 65000 65050 65100 65150 65200 65250 65300 65350 65400 65450 65500 65550 65600 65650 65700 65750 65800 65850 65900 65950 66000 66050 66100 66150 66200 66250 66300 66350 66400 66450 66500 66550 66600 66650 66700 66750 66800 66850 66900 66950 67000 67050 67100 67150 67200 67250 67300 67350 67400 67450 67500 67550 67600 67650 67700 67750 67800 67850 67900 67950 68000 68050 68100 68150 68200 68250 68300 68350 68400 68450 68500 68550 68600 68650 68700 68750 68800 68850 68900 68950 69000 69050 69100 69150 69200 69250 69300 69350 69400 69450 69500 69550 69600 69650 69700 69750 69800 69850 69900 69950 70000 70050 70100 70150 70200 70250 70300 70350 70400 70450 70500 70550 70600 70650 70700 70750 70800 70850 70900 70950 71000 71050 71100 71150 71200 71250 71300 71350 71400 71450 71500 71550 71600 71650 71700 71750 71800 71850 71900 71950 72000 72050 72100 72150 72200 72250 72300 72350 72400 72450 72500 72550 72600 72650 72700 72750 72800 72850 72900 72950 73000 73050 73100 73150 73200 73250 73300 73350 73400 73450 73500 73550 73600 73650 73700 73750 73800 73850 73900 73950 74000 74050 74100 74150 74200 74250 74300 74350 74400 74450 74500 74550 74600 74650 74700 74750 74800 74850 74900 74950 75000 75050 75100 75150 75200 75250 75300 75350 75400 75450 75500 75550 75600 75650 75700 75750 75800 75850 75900 75950 76000 76050 76100 76150 76200 76250 76300 76350 76400 76450 76500 76550 76600 76650 76700 76750 76800 76850 76900 76950 77000 77050 77100 77150 77200 77250 77300 77350 77400 77450 77500 77550 77600 77650 77700 77750 77800 77850 77900 77950 78000 78050 78100 78150 78200 78250 78300 78350 78400 78450 78500 78550 78600 78650 78700 78750 78800 78850 78900 78950 79000 79050 79100 79150 79200 79250 79300 79350 79400 79450 79500 79550 79600 79650 79700 79750 79800 79850 79900 79950 80000 80050 80100 80150 80200 80250 80300 80350 80400 80450 80500 80550 80600 80650 80700 80750 80800 80850 80900 80950 81000 81050 81100 81150 81200 81250 81300 81350 81400 81450 81500 81550 81600 81650 81700 81750 81800 81850 81900 81950 82000 82050 82100 82150 82200 82250 82300 82350 82400 82450 82500 82550 82600 82650 82700 82750 82800 82850 82900 82950 83000 83050 83100 83150 83200 83250 83300 83350 83400 83450 83500 83550 83600 83650 83700 83750 83800 83850 83900 83950 84000 84050 84100 84150 84200 84250 84300 84350 84400 84450 84500 84550 84600 84650 84700 84750 84800 84850 84900 84950 85000 85050 85100 85150 85200 85250 85300 85350 85400 85450 85500 85550 85600 85650 85700 85750 85800 85850 85900 85950 86000 86050 86100 86150 86200 86250 86300 86350 86400 86450 86500 86550 86600 86650 86700 86750 86800 86850 86900 86950 87000 87050 87100 87150 87200 87250 87300 87350 87400 87450 87500 87550 87600 87650 87700 87750 87800 87850 87900 87950 88000 88050 88100 88150 88200 88250 88300 88350 88400 88450 88500 88550 88600 88650 88700 88750 88800 88850 88900 88950 89000 89050 89100 89150 89200 89250 89300 89350 89400 89450 89500 89550 89600 89650 89700 89750 89800 89850 89900 89950 90000 90050 90100 90150 90200 90250 90300 90350 90400 90450 90500 90550 90600 90650 90700 90750 90800 90850 90900 90950 91000 91050 91100 91150 91200 91250 91300 91350 91400 91450 91500 91550 91600 91650 91700 91750 91800 91850 91900 91950 92000 92050 92100 92150 92200 92250 92300 92350 92400 92450 92500 92550 92600 92650 92700 92750 92800 92850 92900 92950 93000 93050 93100 93150 93200 93250 93300 93350 93400 934

BLAST Output: Alignments

```
>ref|NP_940807.1| UG sarcomeric mitochondrial creatine kinase precursor [Mus musculus]
ref|NP_00121124.1| UG sarcomeric mitochondrial creatine kinase precursor [Rattus norvegicus]
Length=419

GENE ID: 76722 Ckm2 | creatine kinase, mitochondrial 2 [Mus musculus]
(Over 10 PubMed links)
```

Score = 511 bits (1316), Expect = 7e-145, Method: Compositional matrix adjust.
 Identities = 239/361 (66%), Positives = 285/361 (78%), Gaps = 2/361 (0%)

Query	Sbjct	Score
Query 14	+ P + YDPL KPI	73
Sbjct 48	FPPSADPYPLRK+	107
Query 74	CVAGVEDWFPVYKLRHNGYDPRVMKHPYD	132
Sbjct	P+I RH GY P XH TDL+ DLDLDPNVLSSRRV	166

Identical match

positive score (conservative)

Negative or zero

gap

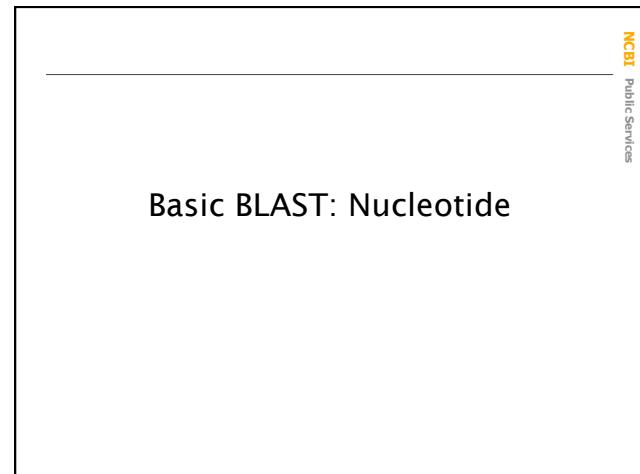
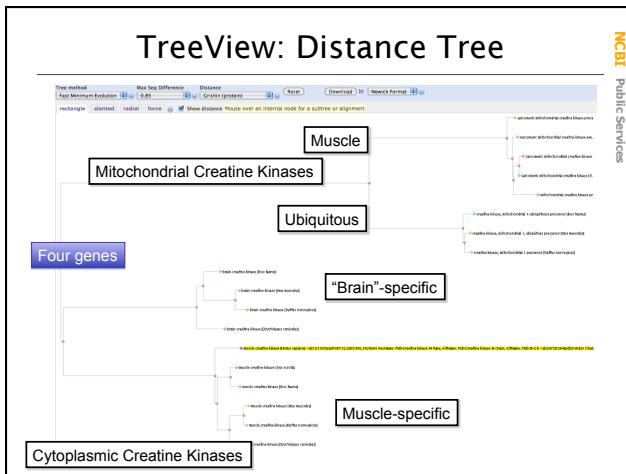
What happens without XP_ filter?	
Sequences producing significant alignments:	Score (Bits) E Value
ref XP_533641_2 PREDICTED similar to Creatine kinase M-type...	.772 0.0
ref XP_537561_2 PREDICTED similar to Creatine kinase B-type...	.653 0.0
ref XP_866559_1 PREDICTED similar to Creatine kinase B-type...	.551 1e-156
ref NP_001138666_1 sarcomeric mitochondrial creatine kinase...	.509 6e-144
ref XP_535451_1 PREDICTED similar to creatine kinase, mitoc...	.503 4e-142
ref XP_860111_1 PREDICTED similar to creatine kinase, mitoc...	.501 2e-141
ref XP_859467_1 PREDICTED similar to creatine kinase, mitoc...	.498 2e-140
ref XP_856432_1 PREDICTED similar to creatine kinase, mitoc...	.497 3e-140
	.433 8e-121
	.412 2e-114
	.358 3e-98
	.318 2e-86
Results filtered for domestic dog proteins.	.239 2e-62
26 additional gene predictions from Dog alone.	.224 6e-58
Many are extra splice variants predicted by Gnomon.	.202 1e-51
	.199 1e-50
	.197 8e-50
	.144 8e-34
	.140 1e-32
	.103 9e-22
	.98.6 4e-20
	.97.8 6e-20
	.95.1 4e-19
	.74.7 6e-13
	.72.8 2e-12
	.31.6 6.1

NCBI Public Services

Other Reports

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#) [TreeView](#) [Tax BLAST](#) [COBALT extension](#)

TaxBLAST: Taxonomy Reports					
Lineage Report					
<i>Rutherford</i> [plasmids]					
<i>Laurasithorax</i> [picisents]					
+ <i>Cervidae</i> [even-toed ungulates]					
+ <i>Bos scoticus</i> [wild boar]	773	2	hits	[even-toed ungulates]	
+ <i>Bos taurus</i> (cow)	768	4	hits	[even-toed ungulates]	
+ <i>Bos taurus</i> (domestic cattle)	768	4	hits	[even-toed ungulates]	
<i>Mus musculus</i> (mouse)	770	0	hits	[rodents]	
<i>Oryctolagus cuniculus</i> (domestic rabbit)	768	4	hits	[rabbits & hares]	
<i>Rattus norvegicus</i> (brown rat)	766	4	hits	[rodents]	
Organism Report		Four genes in each mammal.			
<i>Bos serrota</i> (wild boar, ...)	[even-toed ungulates]		taxid	9823	
refINP_001038036.1	sarcomeric mitochondrial creatine kinase		773	0.0	
refINP_001038036.1	sarcomeric mitochondrial creatine kinase		509	2e-144	
<i>Mus musculus</i> (mouse)	[rodents]		taxid	10990	
refINP_03179.1	muscle creatine kinase		[Mus musculus]	770	0.0
refINP_03179.1	brain creatine kinase		[Mus musculus]	552	0.0
refINP_049080.1	sarcomeric mitochondrial creatine kinase		[Mus musculus]	511	8e-145
refINP_03402.1	creatine kinase, mitochondrial 1 (ubiquit.		[Mus musculus]	304	6e-143
<i>Bos taurus</i> (cow, ...)	[even-toed ungulates]		taxid	9913	
refINP_00107708.1	muscle creatine kinase		[Bos taurus]	768	0.0
refINP_001015613.1	brain creatine kinase		[Bos taurus]	664	0.0
refINP_001015613.1	sarcomeric mitochondrial creatine kinase		[Bos taurus]	550	2e-144
refINP_776700.1	creatine kinase, mitochondrial 1 (ubiquit.		[Bos taurus]	504	7e-143
<i>Oryctolagus cuniculus</i> (domestic rabbit, ...)	[rabbits & hares]		taxid	9986	
refINP_00107708.1	muscle creatine kinase		[Oryctolagus cu...]	768	0.0
refINP_001164561.1	sarcomeric mitochondrial creatine kinase		[Oryctolagus cu...]	550	0.0
refINP_001164561.1	mitochondrial creatine kinase precursor		[Oryctolagus cu...]	509	1e-141
<i>Rattus norvegicus</i> (brown rat, ...)	[rodents]		taxid	9416	
refINP_036667.1	muscle creatine kinase		[Rattus norvegicus]	766	0.0
refINP_001121324.1	creatine kinase		[Rattus norvegicus]	550	0.0
refINP_001121324.1	sarcomeric mitochondrial creatine kinase		[Rattus norvegicus]	511	8e-145
refINP_001015613.1	creatine kinase, mitochondrial 1 precu...		[Rattus norvegicus]	309	4e-143
<i>Canis lupus familiaris</i> (dog)	[canines]		taxid	9415	
refINP_001136851.1	sarcomeric mitochondrial creatine kinase		[Canis lupus famili...	509	2e-144



Universal Form: Nucleotide

The interface includes a toolbar with File, Edit, Help, and Bookmarks. A legend on the right indicates that green arrows point to 'More' and red arrows point to 'Less'.

Nucleotide Results: ALB mRNA

Accession	Description	Max score	Total score	Query coverage	
megablast					
NM_000477.3	Homo sapiens albumin (ALB), mRNA	3995	3995	100%	
NM_001020561.1	Felis catus albumin (Alb), mRNA	1327	1327	93%	
NM_001025203.1	Equus caballus preproalbumin (LOC100034206), mRNA	1315	1919	91%	
NM_001055084.1	Sus scrofa albumin (ALB), mRNA	1827	1827	92%	
disco. megablast					
NM_000477.3	Homo sapiens albumin (ALB), mRNA	3995	3995	100%	
NM_001020561.1	Felis catus albumin (Alb), mRNA	2132	2132	91%	
NM_001025203.1	Equus caballus preproalbumin (LOC100034206), mRNA	2118	2118	84%	
NM_001055084.1	Sus scrofa albumin (ALB), mRNA	2085	2085	92%	
NM_001020773.1	Ovis aries pre-pro serum albumin (LOC445393), mRNA	1397	1397	92%	
NM_001020344.1	Oryctolagus cuniculus serum albumin precursor (LOC100009107), mRNA	1539	1983	87%	
NM_18892.2	Bos taurus albumin (ALB), mRNA	1563	1963	92%	
NM_009546.2	Mus musculus albumin (Alb), mRNA	1524	1524	92%	
NM_134520					
NM_205261	Accession	Description	Max score	Total score	Query coverage
NM_17595.1	NM_000477.3	Homo sapiens albumin (ALB), mRNA	3995	3995	100%
NM_001020561.1	NM_001082053.1	Equus caballus preproalbumin (LOC100034206), mRNA	2132	2132	91%
NM_001025203.1	NM_001009951.1	Felis catus albumin (Alb), mRNA	2118	2118	84%
NM_001055084.1	NM_001020773.1	Sus scrofa albumin (ALB), mRNA	2085	2085	92%
NM_0011332.1	NM_001020344.1	Ovis aries pre-pro serum albumin (LOC445393), mRNA	1539	1983	87%
NM_001020927.1	NM_000477.3	Oryctolagus cuniculus serum albumin precursor (LOC100009107), mRNA	1539	1983	87%
NM_001020344.1	NM_001082053.1	Bos taurus albumin (ALB), mRNA	1563	1963	92%
NM_009546.2	NM_134520	Mus musculus albumin (Alb), mRNA	1534	1534	92%
NM_134520	NM_001020773.1	Rattus norvegicus albumin (ALB), mRNA	1449	1449	91%
NM_17595.1	NM_001020344.1	Bos taurus albumin (ALB), mRNA	127	127	46%
NM_17595.1	NM_134520	Bos taurus conglutinin (C0N3), mRNA	65.5	66.2	4%
NM_001332.1	NM_001020773.1	Homo sapiens afamin (AFM), mRNA	55.4	55.4	2%
NM_001094613.1	NM_001020344.1	Xenopus laevis 68 kDa serum albumin (LOC397731), mRNA	55.4	55.4	7%
NM_001094613.1	NM_001020773.1	Xenopus laevis 74 kDa serum albumin (LOC448844), mRNA	53.6	53.6	8%

Macaque CDC20 Search

GenBank: AB106836.1

Macaque fascicularis testis cDNA clone: QtsA-13692, similar to human CDC20 cell division cycle 20 homolog (S. cerevisiae)(CDC20), mRNA, RefSeq: NM_001255.1

Comments Features Sequence

LOCUS AB106836
DEFINITION Macaque fascicularis testis cDNA clone: QtsA-13692, similar to human CDC20 cell division cycle 20 homolog (S. cerevisiae)(CDC20), mRNA, RefSeq: NM_001255.1

ACCESSION AB106836
VERSION AB106836.1
KEYWORDS FJL;CMNA; oligo capping;
SOURCE Macaque fascicularis (crab-eating macaque);
ORGANISM Bulyarukia Metacercus (Chordata: Primates: Hominoidea: Hominidae: Catarrhini: Hominoidea)

REFERENCE
AUTHORS Osada,N., Mirata,M., Tandjou, S., Sugeno,S., Gojobori,T., & Matsunaga,K.
TITLE Molecular phylogenetic analysis based on comparative analysis between Mol. Biol. Evol. 22 (10):
JOURNAL Mol. Biol. Evol.
PUBLISHED
REFERENCE 2
COUNTRY International consortium
TITLE DNA sequences of macaque evolutionary implications
JOURNAL Mol. Biol. Evol.
PUBLISHED
REFERENCE 3 (bases 1 to 1696)
SEQUENCE

Change Region Shown
Customize View

Sequence Analysis Tools
BLASTN Sequence
Print/Printers

NCBI Public Services

Enter Query Sequence
 BLASTN programs search nucleotide databases using a nucleotide query, maxmimum alignment length = 1000 bp, e-value cutoff = 10.0, word size = 11, mismatch cost = 2, gap cost = 10.

Enter accession number, gi, or FASTA sequence AB106836.1
From:
To:

Or upload file **Browse**

Job Title AB106836 Macaque fascicularis testis cDNA clone.
 Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database Human genomic + transcript Mouse genomic + transcript Other (nr etc.)
 Human genomic plus transcript (Human G+T)

Exclude optional Molecules (*molc*) Entrez nucleotide sample sequences

Entrez Query
Optional Enter an Entrez query to limit search

Sortable Results

Distribution of 23 Blast Hits on the Query Sequence

NT_032977 Homo sapiens chromosome 1 genomic contig, GRCh37 reference, S=215 E=1.5e

Query

Descriptions

Legend for links to other resources: Uniprot Gene Structure Map Viewer

Sequences producing significant alignments
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E – value	Max ident	Links
Transcripts	Homo sapiens cell division cycle 20 homolog (S. cerevisiae) (CD20),	2833	2839	98%	0.0	97%	
Genomic sequences (show first)	Homo sapiens chromosome 3 genomic contig, GRCh37 reference pret	2673	2673	97%	0.0	95%	
NT_000470_19	Homo sapiens chromosome 3 genomic contig, GRCh37 reference pret	2649	2649	97%	0.0	95%	
NW_000185222.1	Homo sapiens chromosome 3 genomic contig, alternate assembly pret	2643	2643	97%	0.0	95%	
NT_032927_3	Homo sapiens chromosome 1 genomic contig, GRCh37 reference pret	511	2853	94%	2e-111	100%	
NW_000185201.2	Homo sapiens chromosome 1 genomic contig, alternate assembly (b)	612	2853	94%	2e-111	100%	

Pseudogene on Chromosome 9

Functional Gene on Chromosome 1

Total Score: All Segments

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
Transcripts						
LOC_0012420	140S ribosomal cell division cycle 20 homolog (S. cerevisiae) (CDC20)	2833	2839	98%	0.0	97%
Genomic sequences [show first]						
NC_001838578.2	Homo sapiens chromosome 1 genomic contig, alternate assembly (b- NT_022977.9	411	2853	94%	2e-111	100%
NT_008470.19	Homo sapiens chromosome 1 genomic contig, GRCh37 reference prf NW_001838222.1	411	2853	94%	2e-111	100%
	Homo sapiens chromosome 9 genomic contig, GRCh37 reference prf Homo sapiens chromosome 9 genomic contig, alternate assembly (b-	2673	2673	97%	0.0	95%
		2643	26	97%	0.0	95%

Functional Gene Now First

Links to Map Viewer

NCBI Public Services

BLAST Formatting Options

NCBI Public Services

Formatting Page (Now on Results)

NCBI Public Services

Reformatted Results

NCBI Public Services

NCBI Public Services

Download Options (Now on Results)

Download					
Alignment Text XML ASN.1 Hit Table(text) Hit Table(csv)	Search Strategies ASN.1	Bioseq ASN.1	PSSM Text	PssmWithParameters ASN.1	
Structured Formats Saved Settings <ul style="list-style-type: none"> Reusable on Web Portable to Standalone 					
PSSM <ul style="list-style-type: none"> Reusable on Web Portable to Standalone 					

The Hit Table

Structured formats: XML and ASN.1	ASN.1
<pre> <Iteration_hits> <Hit> <Hit_num>1</Hit_num> <Hit_id>qj173028 ep P40692 MLH1_HUMAN</Hit_id> <Hit_def> DNA mismatch repair protein MLH1 (MutL homolog 1) </Hit_def> <Hit_accession>P40692</Hit_accession> <Hit_len>756</Hit_len> <Hit_bit_score>1568.9</Hit_bit-score> <Hit_bit-score>1568.9</Hit_bit-score> <Hit_score>1041</Hit_score> <Hit_eval>0</Hit_eval> <Hit_query_from>1</Hit_query-from> <Hit_query_to>756</Hit_query-to> <Hit_bit-from>1</Hit_bit-from> <Hit_bit-to>756</Hit_bit-to> <Hit_query-frame>0</Hit_query-frame> <Hit_bit-frame>0</Hit_bit-frame> <Hit_identity>0</Hit_identity> <Hit_positive>0</Hit_positive> <Hit_gaps>0</Hit_gaps> <Hit_align-len>756</Hit_align-len> </Hit> </Iteration_hits> </pre>	<pre> Seq-align ::= { desc { user { type { str "Hist Seqalign", data { { label str "Hist Seqalign", data bool TRUE } } }, user { type { str "Blast Type", data { { label id 0, data int 0 } } }, user { type { str "BLAST database title", data { { label str "Non-redundant SwissProt" } } } } } } } </pre>

PSSMs: Restart PSI-BLAST

```

FssmWithParameters ::= {
  pssn {
    numRows 28,
    numColumns 157,
    query seq {
      id {
        other {
          accession "YP_385691"
        }
      },
      inst {
        repr raw,
        mol aa,
        length 616
      }
    }
  },
  ASN_1 ScoreMat, Portable
  intermediateData {
    freqTable {
      { 0, 10, 0 },
      { 427396, 10, -7 },
      { 0, 10, 0 },
      { 593523, 10, -8 },
      { 18841, 10, -6 },
      { 231628, 10, -7 },
      { 729652, 10, -8 },
      { 245444, 10, -7 },
      { 622078, 10, -8 },
    }
  }
}

Fssm ::= {
  FSSM_Seq {
    FSSMWithParameters
    FSSM_Seq
  }
}

FSSM_Seq ::= {
  FSSMWithParameters
  FSSM_Seq
}

```

NCBI Public Services

Managing Searches

Recent Results
Saved Strategies

NCBI Public Services

Recent Results

BLAST **BETA**

Links to your unexpired BLAST jobs appear below. [more...](#)

Lookup BLAST Job
RID: Go

Submitted at	RID	Status	Program	Title	Oligonucleotide length	Database	Expires at
03-13 15:01	ZDX95NAK012	Done	tblastn	AA000454.adspin [Smithothrixthus anatinus]	159	wgs	03-15 03:01
03-13 14:57	ZDVWVZVDB012	Done	tblastn	g10837159 (476 letters)	476	wgs	03-15 02:57
03-13 14:35	ZDVB8V8M013	Done	tblastn	(4) - g1t2544239[refNM_001928.2]Homo sapiens...	1173	wgs	03-15 02:35
03-13 14:33	ZDVXKE5013	Done	tblastn	(3) - g1t2544239[refNM_001928.2]Homo sapiens...	1173	wgs	03-15 02:33
03-13 14:29	ZDVB9HA2012	Done	tblastn	(2) - g1t2544239[refNM_001928.2]Homo sapiens...	1173	wgs	03-15 02:29
03-13 14:28	ZDVAJN2H012	Done	tblastn	g1t2544239[refNM_001928.2]Homo sapiens...	1173	wgs	03-15 02:28
03-13 14:21	ZDUWMTE013	Done	tblastn	g1t2544239[refNP_001919.2]complement factor...	253	wgs	03-15 02:21
03-13 12:47	ZDSVAK06013	Done	blastp	protein sequence 10 aa	10	nr	03-15 01:47

Copyright | Disclaimer | Privacy | Accessibility | Contact | Send feedback on new interface

NCBI | BLAST | BLAST+ (beta) | BLASTN | BLASTP | BLAT | BLAT+ (beta) | BLATN | BLATP | BLATX | BLAT+X (beta) | BLATNX | BLATXP | BLATX+ (beta) | BLATNXP | BLATX+P (beta) | BLATNP | BLATX+P+ (beta) | BLATNPP | BLATX+P+P (beta) | BLATNPP

NCBI Public Services

Saved Strategies

Basic Local Alignment Search Tool

NCBI BLAST Saved Strategies

Your Saved Strategies

Enter accession number, g1, or FASTA sequence Clear Query change

From To

Or, upload file Browse... Job Title: Re-run searches to keep up to date

Choose Search Set Database: Non-human, non-mouse ESTs (est_other) Organism: Salmonella (taxd015) Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Enter Query Options: Enter an Entrez query to limit search

BLAST Search database est_other using tblastn (Search translated nucleotide database using a protein query)
Show results in a new window

NCBI Public Services

Genome and Specialized BLAST

Nucleotide Databases: Human and Mouse

Megablast, blastn service

Choose Search Set

Database: Human genomic + transcript Mouse genomic + transcript Others (nr etc.)
 Human genomic plus transcript

- Human and mouse genomic and transcript now default
- Separate sections in output for mRNA and genomic
- Direct links to Map Viewer for genomic sequences

Genome BLAST pages

BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

<input type="checkbox"/> Human	<input type="checkbox"/> Oryza sativa
<input type="checkbox"/> Mouse	<input type="checkbox"/> Gallus gallus
<input type="checkbox"/> Rat	<input type="checkbox"/> Bos taurus
<input type="checkbox"/> Arabidopsis thaliana	<input type="checkbox"/> Pan troglodytes
	<input type="checkbox"/> Danio rerio
	<input type="checkbox"/> Microbes
	<input type="checkbox"/> Drosophila melanogaster
	<input type="checkbox"/> Apis mellifera

Map Viewer Homepage

NCBI Public Services

Search: Monocots

Tools Legend:

- Search or Browse the Genome
- BLAST
- Genome Resources page

New:

Annotation update plans for Human
Genome build X
NCBI is planning to update its annotation of the human genome. [more](#) Show all

Related Resources:

- NCBI Home
- NCBI Web Search
- NCBI Site Map
- Genomic Biology
- Taxonomy
- Entrez (Global Query)
- BLAST
- Map Viewer FTP

Small Genomes:

- Bacteria
- Organellas
- Viruses

Monocots (13) (21)

Scientific name	Common name	Build	Tools
<i>Arachis hypogaea</i>	peanut	Build 1.0	<input type="radio"/> <input type="radio"/>
<i>Beta vulgaris</i>	beet	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Brassica juncea</i>	brown mustard	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Brassica napus</i>	rapeseed	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Brassica nigra</i>	black mustard	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Brassica oleracea</i>		Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Brassica rapa</i>	field mustard	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Capsicum annuum</i>	pepper	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Glycine max</i>	soybean	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Lotus japonicus</i>	birdfoot trefoil	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Manihot esculenta</i>	cassava	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Medicago sativa</i>	alfalfa	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Phaseolus vulgaris</i>	kidney bean	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Polygonum perfoliatum</i>	popcorn	Build 1.1	<input type="radio"/> <input type="radio"/>
<i>Prunus dulcis</i>	sweet almond	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Solanum lycopersicum</i>	nightshade	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Solanum melongena</i>	tomato	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Solanum peruviana</i>	eggplant	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Pennisetum glaucum</i>	Pennisetum tomato	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Theobroma cacao</i>	cocoa	Build 7.0	<input type="radio"/> <input type="radio"/>
<i>Vigna radiata</i>	mungbean	Build 7.0	<input type="radio"/> <input type="radio"/>

Poplar Genome BLAST

NCBI Public Services

BLAST Overview FAQs News Manual References Retrieve results Genome Project

Enter an accession, ID, or a sequence in FASTA format:
>gi|6911075|gb|1AP19893|1-114154527 1 chyle receptor
MEACNCIEPQFPAELLMETMKTQTSDFYI1ALATFSPLPLILTVYKEAVPFVWVWVQGQAFVLC
TSHLRLVLL
ETYGGRVNVMLTRHITSLPQPTFLTTVVLIGGTALKALCALWNTT1T1GLQLQSTTLPGQNPQVWT
LPVIVQVTSSEKAKLTKSPHPPVARMPLGKHMPCPEVAVPVPLHLHNPQINQDPFLSTKTA
LSQGQXPRVHLLTVEVAVQVAYASLXAALEERHPAFSLDQHQLXQDLPFFKTAIP
C or, choose a file to upload:

Set subsequences: (optional)
From: To:

Database: genomic (reference only) 45688 sequences

Program: BLASTN: Compare a protein sequence against a nucleotide database
 megabLAST: Compare highly related nucleotide sequences
 cross-species megabLAST: Compare nucleotide sequences for other species to this genome
 ESTs: Compare ESTs against a nucleotide database
 Traces-other*: WGS contigs: Compare a nucleotide database against a protein database
 BLASTP: Compare protein sequences
 TBLASTN: Compare a protein sequence against a nucleotide database
 Advanced options:

Begin Search

NCBI
Public Services

Genomic Context of BLAST Hits

Populus trichocarpa (poplar) genome view

Build 1.1 statistics

Linkage Group	Hits (Top Row)	Hits (Bottom Row)
LGI	1	1
LGII	4	1
LGIII	3	3
LGIV	1	4
LGV	1	1
LGVI	1	1
LGVII	6	1
LGVIII	1	1
LGIX	4	45
LGX	1	71
LGIX not placed	0	0

BLAST search *Populus trichocarpa* sequences

NCBI Public Services

Specialized BLAST Pages

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with [Primer-BLAST](#)
- Search [trace archives](#)
- Find [conserved domains](#) in your sequence (cds)
- Find sequences with similar [conserved domain architecture](#) (cdart)
- Search sequences that have [gene expression profiles](#) (GEO)
- Search [immunoglobulins](#) (IgBLAST)
- Search for [SNPs](#) (snp)
- Screen sequence for [vector contamination](#) (vecscren)
- [Align](#) two (or more) sequences using BLAST (bl2seq)
- Search [protein](#) or [nucleotide](#) targets in PubChem BioAssay
- Search [SRA transcript libraries](#)
- Constraint Based Protein [Multiple Alignment Tool](#)

BLAST extensions and improvements

- PrimerBlast – primer designer / specificity checker
- COBALT – Protein Multiple Alignment tool
- Integration / expansion of BLAST 2 Sequences

NCBI
Public Services

Primer BLAST from Sequence Record

NCBI Reference Sequence NM_001077.2

Homo sapiens UDP glucuronosyltransferase 2 family, polypeptide B17 (UGT2B17), mRNA

[Comment](#) [Features](#) [Sequence](#)

LOCUS NM_001077 2099 bp mRNA linear PRI [protein](#)

DEFINITION Homo sapiens UDP glucuronosyltransferase 2 family, polypeptide B17 (UGT2B17), mRNA.

ACCESSION NM_001077 XM_936292

VERSION NM_001077.2 GI:52651461

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Homoplacini; Catarrhini; Homidae; Homo.

REFERENCE UGTR2B17.1 (2009)

AUTHORS Strahak,E., Sottas,P.E., Schweizer,C., Saugy,M., Dvorak,J., and Saudan,C.

TITLE Steroid profiles of professional soccer players: an international comparative study

JOURNAL Br J Sports Med 43 (14), 1126-1130 (2009)

PIPERED

REMARK GeneRIF: Observational study of gene-disease association. (HUGO Navigator)

REFERENCE UGTR2B17.1 (2009)

AUTHORS Ahn,J., Schumacher,F.R., Berndt,B.I., Pfeiffer,R., Albares,D., Andriole,G.L., Ardanaz,E., Boeing,H., Bueno-de-Mesquita,B., Chanock,S.J., Clavel-Chapelon,F., Diver,W.R., Fajgeles,N.H., Gao,Y., Giovannucci,E., Givens,I., Groves,T., Harnack,L., Hoover,R.N., Kolonel,L.N., Kraft,P., Ma,J., Le Marchand,L., Overvad,K., Palli,D., Stattin,P., Stampfer,M., Stram,D.O., Thomas,G., Thun,M.J., Trichopoulou,A., Virtamo,J.,

Sequence Analysis Tools

- BLAST Sequence
- Customize View
- Pick Primer

Articles about the UGT2B17 gene

- Quantitative trait loci prediction of circulating sex steroid hormone levels [PubMed:20086000]
- Copynumber variations (CNVs) of the human sex steroid metabolite (Hum Mut, 2009)
- Characterization of UGTs active against SARA and its inhibitor (Cancer Res, 2008)

[» See all](#)

RefSeq Protein Product

See the reference protein sequence for UDP glucuronosyltransferase 2 family, polypeptide B17 precursor (NP_001077.1)

More about the UGT2B17 gene

UGT2B17 belongs to the family of UDP glucuronosyltransferases (UGTs), EC 2.4.1.17, enzymes that catalyze the transfer of glucuronic acid to

NCBI
Public Services

Primer BLAST: Template and Primers

NCBI
Public Services

PCR Template [Reset page](#) [Save search parameters](#)

Enter accession, gI, or FASTA sequence (A refseq record is preferred) [Clear](#)

NR_001077.2

Range

Forward primer To

Reverse primer

Or, upload FASTA file [Browse...](#)

Primer Parameters

Use my own forward primer (5'→3' on plus strand)
Use my own reverse primer (5'→3' on minus strand)

PCR product size 100 1000

of primers to return 10

Min Opt Max Max Tm difference

Primer melting temperatures (T_m) 57.0 60.0 63.0 3

Please note the recent change in default Tm calculation [?](#)

Primer BLAST: specificity params

Exon/intron selection

A refseq mRNA sequence as PCR template input is required for options in the section [?](#)

Exon junction span

No preference

Exon junction match

Exon at 5' side No preference

Minimal number of exons 4 Primer must span an exon-exon junction

Primer may not span an exon-exon junction

Primer must be separated by at least one intron on the corresponding genomic DNA

Intron inclusion

Min 1000 Max 1000000

Get Primers

Primer Pair Specificity Checking Parameters

Enable search for primer pairs specific to the intended PCR template [?](#)

Organism

9606 [Organism-specific search](#)

Enter an organism name, taxonomy id or select from the suggestion list as you type. [?](#)

Add more organisms [?](#)

Database

RefSeq RNA (refseq_rna)

At least 2 total RefSeq RNA (refseq_rna)

at least 2 unique RefSeq RNA (refseq_rna)

Primer specificity stringency

Genome (reference assembly from selected organisms)

Genome (chromosomes from all organisms)

Allow primer to ampl.

Minimally product size deviation

Splice variant handling

NCBI
Public Services

Primer Results

NCBI Public Services

Primer pair 1	Sequence (5'->3')	Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	AAGCACCTTCGGTCGAGC	Plus	20	1468	1485	59.97	65.00%
Reverse primer	GGAAGCTCAGTAACCTTGTGGGGT	Minus	27	1928	1902	59.17	48.15%
Product length	463						

Products on intended target
NM_001077.2 Homo sapiens UDP glucuronosyltransferase 2 family, polypeptide B17 (UGT2B17), mRNA

product length = 463
Forward primer 1 AAGCACCTTCGGTCGAGC 20
Template 1468 1485
Reverse primer 1 GGAAGCTCAGTAACCTTGTGGGGT 27
Template 1928 1902

BLAST 2 Sequences

NCBI Public Services

Enter accession number, gi, or FASTA sequence

From:
To:

Filters and Masking

Filter: Low complexity regions Species-specific repeats for: Human

Mask: Mask for lookup table only Mask lower case letters

Region between UGT2B15 and TMRSS11E in primary reference and alternate locus

Program Selection

Optimize for: Highly similar sequences (megablast) More dissimilar sequences (discontiguous megablast) Somewhat similar sequences (blastn)

Choose a BLAST algorithm: Megablast BLASTN BLASTP BLASTX TBLASTN TBLASTX

BLAST Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

Algorithm parameters

Alignment of reference and null-allele

NCBI Public Services

Color key for alignment scores

Null Allele (Alt. Locus)

Primary Reference

COBALT Extension of BLAST

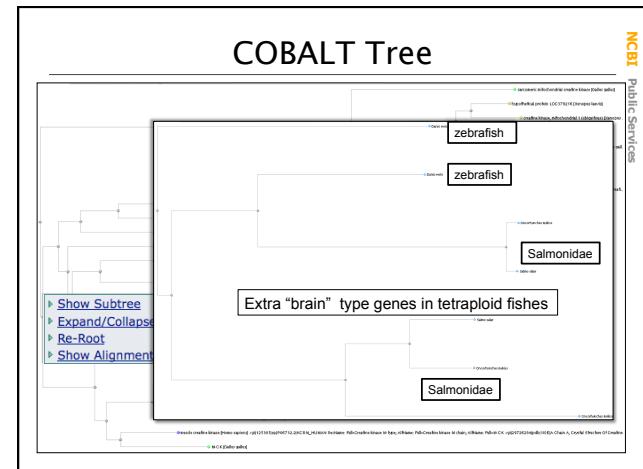
NCBI Public Services

Distribution of 27 Blast Hits on the Query Sequence NP_990838.M-CK [Gallus gallus] S=728 E=0

Color key for alignment scores

Search Summary [Taxonomy reports]	Distance tree of results	Multiple alignment
NP_990838.M-CK [Gallus gallus]	1	1
Mammalia	1	1
Primates	1	1
Hominidae	1	1
Homo	1	1
Homo sapiens	1	1
Muscle creatine kinase [Xenopus laevis]	728	0.0
Muscle creatine kinase [Xenopus (Silurana) tropicalis]	324	0.1
Muscle creatine kinase [Danio rerio]	710	0.0
Creatine kinase-2 [Salmo salar]	706	0.0
Creatine kinase [Xenopus (Silurana) tropicalis]	682	0.0
Creatine kinase [Danio rerio]	681	0.0
Creatine kinase [Danio rerio]	680	0.0
Creatine kinase [Danio rerio]	680	0.0
Creatine kinase [Danio rerio]	679	0.0
Creatine kinase [Danio rerio]	678	0.0
Creatine kinase [Danio rerio]	672	0.0
Creatine kinase [Danio rerio]	671	0.0
Creatine kinase [Danio rerio]	668	0.0
Creatine kinase [Danio rerio]	664	0.0
Creatine kinase b-type [Salmo salar]	662	0.0
Creatine kinase [Oncorhynchus mykiss]	655	0.1
Creatine kinase [Oncorhynchus mykiss]	537	1e-149
Creatine kinase, mitochondrial [Xenopus (...)]	526	2e-149
Creatine kinase, mitochondrial [Xenopus (...)]	525	4e-149
Creatine kinase, mitochondrial [Danio rerio]	523	2e-149
Creatine kinase, mitochondrial [Danio rerio]	522	2e-149
Creatine kinase, mitochondrial [Danio rerio]	518	3e-145
Sarcoplasmic reticulum creatine kinase [Danio rerio]	516	1e-146
Sarcoplasmic reticulum creatine kinase [Danio rerio]	509	2e-144
Hypothetical protein LOC563955 [Danio rerio]	278	1e-14
Hypothetical protein LOC327594 [Danio rerio]	265	4e-71

Lower vertebrate creatine kinases



The screenshot shows the COBALT Interface homepage. At the top, there's a navigation bar with links for Home, Recent Results, and a search bar. The main title "COBALT Interface" is prominently displayed. Below it, the subtitle "Constraint-based Multiple Alignment Tool" is shown. To the right, there are links for "My NCBI" (with a dropdown menu for Sign In and Register), a search bar, and a "Public Services" link. A large orange banner on the right side of the page says "NCBI". The main content area has a blue header "Enter Query Sequences" with a "Reset page" link. It contains a text input field with placeholder text "Enter at least 2 protein accessions, gis, or FASTA sequences" and a "Clear" button. Below this is another text input field labeled "Or, upload FASTA file" with a "Browse" button. There's also a "Job Title" input field. At the bottom, there are buttons for "Align" and "Advanced parameters", and a checkbox for "Show results in a new window".

The screenshot shows the NCBI News homepage with a blue header bar. The title 'Keeping up with what's new' is displayed prominently. Below it, the 'NCBI News' logo is shown with the text 'National Center for Biotechnology Information | National Library of Medicine | National Institutes of Health' and 'Department of Health and Human Services'. A search bar at the top right includes dropdown menus for 'Within' (set to 'This book'), 'All books', and 'PubMed', with a 'Go' button. Below the search bar, a breadcrumb trail reads 'NCBI > Bookshelf > NCBI News >'. The main content area features a large box titled 'NCBI News on Bookshelf' containing the URL 'www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=newsnncbi'. To the left, there is a sidebar with links to news from August 2009, a featured resource about biological pathways, and various news databases and tools. A large list of news items from 2008 is also present. On the far right, a vertical sidebar labeled 'Public Services' lists several services.

Getting Help

NCBI Public Services

BLAST Home Recent Results Saved Strategies Help

NCBI/BLAST/Help

Browse BLAST documentation.

Getting Started

BLAST short course
BLAST program selection guide
BLAST interface description

About BLAST

Frequently Asked Questions
NCBI Handbook: BLAST
The Statistics of Sequence Similarity Scores
NAR 2004 Web server issue
NAR 2006 Web server issue
NAR 2008 Web server issue
BLAST glossary
References

Getting Help

Email blast-help
Mailing list

Other BLAST information

Download BLAST Software and Databases
Developer information

Service Addresses

NCBI Public Services

• **General Help** info@ncbi.nlm.nih.gov

• **BLAST** blast-help@ncbi.nlm.nih.gov

Telephone support: 301-496-2475